DOCUMENT RESUME

ED 359 866 HE 026 555

AUTHOR Moore, Gordon; And Others
TITLE A Randomized Controlled Trial Evaluating the Impact of the New Pathway Curriculum at Harvard Medical School. Final Report.
INSTITUTION Harvard Univ., Cambridge, Mass. Medical School.
SPONS AGENCY Fund for the Improvement of Postsecondary Education (ED), Washington, DC.
PUB DATE 29 Nov 90
CONTRACT G008730472-89
NOTE 43p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; Career Choice; Comparative Analysis; Curriculum Development; *Experimental Curriculum; Higher Education; Interdisciplinary Approach; Medical Education; *Medical Schools; *Medical Students; Mentors; *Outcomes of Education; *Program Evaluation; Student Centered Curriculum
IDENTIFIERS *Harvard University Medical School MA

ABSTRACT

In 1985 and 1986, two randomly selected groups of Harvard Medical School students entered either the traditional program or the New Pathway, a redesigned medical curriculum with a 3-year longitudinal course structure involving the same small group of students and teachers. The program features the interweaving of material from other disciplines; emphasis on self-reflection; a mentoring relationship between students and teachers within the small groups; and opportunities to discuss and reflect on experiences that occurred during rotations. An evaluation of the program's effects involved analysis of career preferences and educational experiences, observed information on student performance, self-report data, and test data from standardized patient and National Board scores. These results showed that the New Pathway students preferred a student-directed environment, studied differently, and demonstrated a stronger orientation towards "deep" learning. The students in the new curriculum were more humanistic in orientation, possessed better relational skills, and had superior knowledge of social and behavioral science. In addition they were more challenged, better known to the faculty, and somewhat more anxious. There were no differences in biomedical cognitive performance measures or career choices. Includes appendixes detailing evaluation instruments, giving summary tables of learning and psychosocial behaviors, and listing 25 references. (JB)

FINAL REPORT TO THE
FUND FOR THE IMPROVEMENT OF POST-SECONDARY EDUCATION
11/28/90

A RANDOMIZED CONTROLLED TRIAL EVALUATING THE IMPACT OF THE NEW PATHWAY CURRICULUM AT HARVARD MEDICAL SCHOOL

Gordon Moore, Susan Block, and Rudolph Mitchell

Cover Sheet

Grantee Organization:

President and Fellows of Harvard College
Office of Sponsored Research
Holyoke Center, 4th floor
Harvard University
Cambridge, MA 02138

Grant Number:

G008730472-89

Project Dates:

Starting Date: September 1, 1987
Ending Date: August 31, 1990
Number of Months: 36

Project Director:

Gordon T. Moore MD, MPH
Office for Educational Development
Medical Education Center
260 Longwood Avenue
Harvard Medical School
Boston, MA 02115

PIPSE Program Officer:

Constance Cook

Grant Award:

Year 1: $123,221
Year 2: $119,770
Year 3: $118,902
Total: $361,893

ABSTRACT

In 1985 and 1986, two randomly selected groups of Harvard Medical School (HMS) students entered either the traditional program or a radically redesigned medical curriculum (The New Pathway) that was built upon many of the suggestions of national committees and the work of adult learning theorists. We conducted an evaluation that showed that the students' learning methods reflected the type of educational process used in each curriculum: the New Pathway students preferred a student-directed environment, studied more differently, and demonstrated a stronger orientation towards "deep" learning. The students in the new curriculum were significantly more humanistic in orientation, possessed better relational skills, and had superior knowledge of social and behavioral science than the control group in the traditional curriculum. In addition, they were more challenged, were better known to the faculty, and were somewhat more anxious. There were no differences in biomedical cognitive performance measures or career choices. The results of this study should encourage other schools to consider such a curriculum.

Name: Gordon T. Moore MD, MPH

Address: Teaching Center
Harvard Community Health Plan
Two Fenway Plaza
Boston, MA 02215

Telephone number: 617-421-2747

PROJECT REPORT:

PROJECT OVERVIEW:

Over the past fifty years, many national committees have written about the need for reform in medical education (Rappelye, 1932; Institute of Medicine, 1983; AAMC,1984). Despite their calls for learner-centered approaches that decrease the passive, lecture-dominated curricula prevalent in the US, few schools have embarked on change. McMaster University was the first (in 1970) to base their entire medical school curriculum on the principles of adult learning. By 1985, fifteen years after the initiation of these efforts, only a handful of schools had furthered the movement towards problem-based learning.

PURPOSE:

We believed that the absence of methodologically sound evaluations of these educational reforms had slowed their extension to other schools. Even today, although some studies have looked at specific issues or questions about student-centered learning, no adequate evaluation of the strengths and weaknesses of learner-centered medical education has been carried out using a randomized controlled trial methodology. Thus schools considering curriculum change have been unable to assess whether important differences truly exist between students from a learner-centered curriculum and similar students from traditional programs. This study reports such an effort.

BACKGROUND AND ORIGINS:

Planning for the Harvard New Pathway (NP) curriculum reform was initiated in 1982 by Dean Daniel C. Tosteson. This revised curriculum was designed and run as a separate track, given in parallel with the traditional medical curriculum at Harvard. The educational philosophy that guided the design and implementation of this program was based on the notion that learning is facilitated when knowledge, attitudes, and skills are taught together in relation to real clinical problems. Our goal was to teach students an integrated approach to thinking about patients that encompassed both biomedical and psychosocial phenomena. Basic science teaching emphasized active, problem-based, student-directed learning in small groups in contrast to the passive, large group lecture method used at most medical schools and in the traditional curriculum at Harvard.

Students' early clinical experiences were complemented by exposure to the social and behavioral sciences and the medical humanities, using a small student-faculty learning group oriented around the theme of the doctor-patient relationship. Educational offerings were consciously designed to shape the values and attitudes of students in directions identified by the faculty as consistent with a humane practice. This course was unique in

several features:

- a 3-year longitudinal course structure involving the same small group of students and teachers;
- interweaving of material from the social and behavioral sciences, ethics, health promotion and disease prevention, and the humanities with teaching about clinical skills;
- continuing emphasis on self-reflection and personal introspection within the small group;
- a "mentoring" relationship between students and teachers within the small groups;
- opportunities in small groups to discuss and reflect on experiences that occur during clinical rotations.

The clinical years of medical school, which had been relatively untouched in educational reforms at other schools, were redesigned to give greater attention to problem-solving and reflective thinking. Continuing basic science experiences were emphasized in the clinical years to stimulate student interest in exploring the scientific issues that are the basis for the practice of medicine.

PROJECT DESCRIPTION:

In academic years 1985 and 1986, first year students, randomly selected from a group of volunteer medical students, entered into this new program. The cohort of unselected volunteers went through the unchanged traditional curriculum along with the remaining students who had chosen not to apply for the new curriculum. Thus, the study population consisted of three groups of students from two classes: those who applied to be in the new curriculum and randomly won entry to the New Pathway (NP); the unselected students who became the control group (C); and those students who had not applied to be in the new curriculum and went through the traditional program (TP), serving throughout the study as a comparison group.

We compared the groups of students to determine the differences in their response to their educational experiences and to assess whether the two curricula were associated with any differences in outcomes. The study covered all four years of medical school. Data of each of the following types were collected (see Appendix for list of all instruments): 1) life data: information about students' life experiences, choices and behaviors (e.g. career preferences, choices; educational experiences); 2) observer data: information based on observation of student performance at particular tasks (e.g. clerkship evaluations, observations of student interactions with patients); 3) self-report data: based on students descriptions of themselves, their reactions, and their attitudes and behaviors; and 4) test data: student responses to a standard and controlled stimulus (e.g. standardized patients, National Board scores).

The assessment was designed to explore as many areas affected by the curriculum as was feasible. For the proposed study, we

grouped the areas of investigation into six categories: (1) entry characteristics and study participation; (2) the cognitive domain: factual knowledge and learning behaviors; (3) clinical performance including reasoning and problem solving; (4) psychosocial orientation; (5) students' personal experience of the old and new curriculum; and (6) career choice.

Each of the six evaluation categories examined by the study was analyzed using appropriate statistics to make comparisons at specific points in time between the two curricula. The three study groups - New Pathway (NP), Non-New Pathway Control (C), and Traditional Pathway (TP) were compared in all analyses. In addition, trends over time for each of the study groups were examined looking for significant shifts and changes within and among the groups.

In this report, we concentrate on the findings for the NP and C groups; since both these groups wanted the NP, these comparisons permit us to distinguish self-selection biases from curricular effects. T-tests and chi square statistic were used where appropriate, with small number corrections. Factor analysis was carried out on some data sets for the purpose of data reduction and conceptual clarification. However, when this technique did not yield meaningful information, it will not be reported. In general, two-tailed tests were used to ascertain significance when we were uncertain of the effect of the educational intervention. However, when literature and the experience of educational experts clearly indicated that the effect of the educational design was unimodal and where we had explicit hypotheses in advance, we utilized one-tailed tests of significance. This will be indicated in the body of the report and in all tables.

PROJECT RESULTS

1. ENTRY CHARACTERISTICS AND STUDY PARTICIPATION

Two hundred and ninety-seven students entered the HMS classes of 1989 and 1990. Of the 123 who applied for the NP, 62 were randomized into the new curriculum (the NP group) and 61 (the C group) remained in the traditional, along with 174 students who had not requested the NP in the first place. Twenty-six percent of the first cohort took time off, mostly after the preclinical years; by the end of our study, however, 86 percent of this class had graduated. Fifty percent of the second class took time off and most of these could not be included in the fourth year evaluations. There were no significant differences in proportions of students taking time off in NP, C, and T groups.

All students were subjected to extensive demographic and informational surveys at the beginning of their first medical school year. These surveys provided baseline data (although students already knew which program they were in) with which the initial comparability of the three groups could be determined. The NP and C groups were very similar at entry with a few exceptions. The study groups did not differ significantly in age,

sex, or parents' educational status. Minority status, available only for the class of 1989, was similar between groups. The groups did not differ significantly on any other demographic measures. The students in the NP did not differ from the C group in college grade point average, research experience, or Medical College Admission Test scores.

We measured a number of other characteristics both to determine if students in the three study groups differed and to secure baselines against which to measure changes during medical school. Many characteristics were the same for NP and C groups. These included:

- career preference;
- students' orientation towards study for "deep" learning (for understanding and meaning) or superficial learning directed towards passing exams (Instrument: Short inventory of approaches to studying:Entwhistle);
- the student's sensitivity to the complexities of moral issues (the Defining Issues Test);
- orientation towards affiliation, desire for cognitive structure, exhibitionism in class, and nurturing behavior toward others (Personality research form:by Douglas Jackson.

The NP and C groups differed (at the p=.05 level) in some specific areas on several measures. Expectations about medicine, although similar, showed the NP students to be slightly more hopeful that they would have long-term relationships with patients while the C group were more optimistic that they would have control over their working hours, have patients with treatable disease, and have available research funding. By the beginning of school, students selected for the NP expressed a greater preference for discovery style learning and disliked passive or teacher-directed learning. They preferred curriculum innovation and were less interested in clarity of what was taught. NP students were somewhat more positively oriented to preventive medicine than students in the C group.

In summary, the NP and C groups were largely similar. Some statistically significant differences in several test items distinguished the groups, but these differences were small. The differences could have occurred by chance variation, arisen from failure of randomization to create truly equivalent groups, or could have reflected an early student adaptation to the methods and expectations of the curriculum to which they had been assigned.

2. COGNITIVE BEHAVIOR: CONTENT KNOWLEDGE AND LEARNING BEHAVIORS

The goal of program evaluation in the cognitive domain was to investigate the effects of the NP and TP curricula on the knowledge and learning behavior of their respective students. From the perspective of cognitive psychology, the essence of the NP is its learning environment, which consists of the case-based method operating within the framework of student-directed tutorials. At

HMS, as at other medical schools, the faculty extensively debated the ability of the tutorial milieu to provide medical students with a learning experience that would enable them to reach the level of understanding expected by the faculty. Critics of the NP had voiced concerns that student-directed tutorials would weaken basic science knowledge. The first priority for the program evaluation in the cognitive domain was to address this issue.

A. CONTENT KNOWLEDGE

The debate about content knowledge was resolved, in part, by having all students take Part I of the National Board of Medical Examiners (NBME). While the National Boards has distinct limitations as a summative assessment tool, it does represent a well-standardized and accepted measure of minimum student factual knowledge. We compared numeric scores of the two groups on sub-test and overall scores. In addition to the NBME, part I, we administered an experimental component of the NBME to the class of 1990. This test (the pattern recognition task) was undergoing evaluation for use in future versions of part II of the NBME. Third, we administered a biochemistry review exam and a test of preventive medicine knowledge to the class of 1989 in the fourth year. Finally, we compared third year medicine clerkship evaluations for evidence of differences in knowledge base.

Except for the anatomy subscores for 1989 students and the behavioral science subscores for both classes, no differences existed among the groups in content knowledge as measured by the NBME. The 1989 NP group scored lower ($p=.05$) on anatomy and higher on behavioral science ($p=.05$) than the C group students who had undertaken the traditional curriculum; the 1990 NP and C group scores were almost identical except for behavioral science, which was higher ($p=.09$) for the NP group. The total NBME score difference between NP and C groups was less than one point.

Almost none of the other measures showed any differences between NP and C groups. Mean scores were very close among the three groups on the Pattern Recognition Task, which was administered to approximately 90 students as part of the class of '90 end-of-second year assessment: NP (79%), TP (81%), and C (81%). The preventive medicine test scores were higher for the 1990 NP than C students ($p=.09$ two-tailed and .04 one-tailed).No significant differences in knowledge between groups could be demonstrated from third year medicine clerkship evaluations or the biochemistry examination. We concluded that the biomedical knowledge bases of students in the NP and traditional curricula, as evaluated by our measures, are generally comparable.

B. LEARNING BEHAVIOR

The second area of investigation focused on learning. We explored two questions: what type of learning environment do students prefer and how do students learn? We found major

differences between NP and traditional program students in both areas (see appendix, Table III).

Preferred Learning Environment: Students in the study groups from both classes differed in their preferences about learning environment as measured during their preclinical years on the Preferred Learning Style Index (PLSI) and the Work Environment Scale (WES). Group means of the PLSI total scores on the first and second year administration of the PLSI were significantly higher ($p=.05$) for the NP than for the TP and C groups. These results reflect an initial preference among NP students for an unstructured learning environment; this preference remained strong through the first year. In contrast, the TP students expressed an initial desire for a structured learning environment that lessened slightly at the beginning of the second year but remained relatively strong.

The PLSI mean scores show an important behavior pattern of the control group. We initially expected, and found, the C group to prefer a learning environment similar to that of the NP group; the C mean score initially differed from the TP mean score at a statistically significant level ($p=.05$). By the start of the second year, the C group mean score was lower than it had been at the first year orientation and had dropped almost to the mean of the TP group. Although the C group initially preferred the unstructured environment, they came over time to hold a preference for a structured environment almost as strong as that of the TP students. We interpret this pattern as a reflection of the students' adaptation to their learning environment.

The Work Environment Scale showed a similar pattern. The NP group at the start of medical school was significantly higher than the C group in their preference for faculty support ($p=.03$) and educational innovation ($p=.01$) and lower in their preference for clarity ($p=.05$). By the end of their first, and through their second years, the NP group had diverged strikingly from the C group in their learning environment preferences. They preferred involvement ($p<.001$), faculty support ($p<.001$), autonomy in learning ($p<.001$), innovation ($p<.001$), and a high work effort environment ($p=.01$). The C and TP groups were significantly higher than the NP in their preference for clarity in their learning setting ($p<.001$).

Approaches to Learning: How students learn under differing educational methods has been of critical importance to educational theorists. The proponents of small group, problem-based learning have long argued that this educational method leads to a deeper and more active form of learning than that acquired by students attending lectures. Although these ideas have their origins in Jean Piaget's early work on conceptualization as active learning, there is a paucity of studies on the effects of problem-based education on learning behavior, especially in medical education research. It is in this context that learning behavior emerged as our second area of investigation.

Finding appropriate measures of learning behavior proved difficult. Within the discipline of cognitive psychology, there are

a number of competing schools of thought, each of which has its own model of learning behavior. A review of the literature on cognitive behavior quickly revealed too many models and instruments (Curry, 1983; Newble and Entwistle, 1986) and too little agreement about their validity and effectiveness among researchers and educators (Freedman and Stumpf, 1980; Fox, 1984). To compensate for these problems and to broaden the scope of our inquiry, we developed and administered a specially constructed general survey of student learning methods, as well as using several other instruments to assess learning styles and preferences.

Our survey -- which we called the Cognitive Behavior Survey (Cog Survey) -- was constructed as a vehicle for exploring the development of students' learning behavior in an academic medical environment. It was based on the theories of cognitive behavior that we believed were applicable to the study of learning of medical students; these included the roles of comprehension (Rumelhardt, 1980; Smith, 1975), the use of analogies and metaphors (Gallagher, 1978), active learning (Barrows and Tamblyn, 1980; Piaget, 1971), visualization (Kosslyn, 1980), model building (Gentner and Stevens, 1983; Johnson-Laird, 1983), and learning styles (Entwistle, 1981) in the cognitive behavior of medical students.

We administered the Cog Survey twice during the evaluation -- at the end of the second and fourth years. Unfortunately, return rates for the survey were adequate to support analysis only for the class of 1989 at the end of its second year. The data, which were consistent with and largely reinforced the learning environment preferences reported above, added an important new piece of information regarding the extent to which the students' learning style emphasized memorization or conceptualization. Memorization refers to cognitive processes that center on rote learning (e.g., drill and repetition, use of flash cards). Conceptualization relates to the use of active processes in learning, such as the construction of mental models, the visualization of physiological processes, or the summarization of complex material.

Student responses to the Cog Survey administered to the class of 1989 at the end of their second year showed that students utilized different methods for learning. The NP students had the lowest mean scores on the memorization scale and the highest on the conceptualization scale. On memorization, the NP students were significantly lower ($p=.02$ and .01 respectively) than either C and TP groups. They were higher than C ($p=.02$) and TP ($p=.13$) on conceptualization as a learning method. These findings for the class of 1989 suggest that the problem-based curriculum of the NP encouraged a learning style among its students that consists of greater conceptualization and less memorization than that found in the learning behavior of the students in the traditional program. The 1989 NP and C groups were indistinguishable at entry to medical school on a similar measure of orientation towards studying -- deep versus superficial learning (Entwistle); consequently, we believe that the differences at the end of the second year can justifiably be ascribed to a curriculum effect.

As mentioned, we do not have sufficient data from the 1990 students to either confirm or refute these results. We distributed the Cog Survey to them during the time period that coincided with their end-of-second year assessment, but only 23 TP students, 22 NP students, and 7 C students (less than a third of those eligible) completed it; returns at the end of year four were even lower. We did not see significant differences between these groups but felt that we could not justify any conclusions on this small and probably biased sample.

However, the class of 1990 responses on the Measure of Intellectual Development (MID) did show differences that indirectly support our conclusion of the effect of the NP curriculum on the learning behavior of its students. The MID was administered to 1990 students as part of their end-of-second year assessment. The MID survey asks students to describe their best learning experience in an essay, which is scored based on the students' stage of development within the Perry schema. The NP mean score was the highest among the three groups; a one-way ANOVA was significant at the p=.05 level. Further analysis by the Scheffe method identified the difference between NP and TP group means as being significant at the p=.05 level. A t-test between the NP and TP group means was significant at the p=.01 level, although the NP and C group difference did not achieve statistical significance. In the context of their descriptive essays, the NP students were judged to be at a point further along on the Perry developmental scale than their TP counterparts. These results suggest that NP students view knowledge in a more relative context, see professors more as resources and less as authority figures, and judge new information more critically than their TP peers. Such behaviors are viewed as products of a conceptual learning style, and thus the results provide indirect evidence that the class of 1990 also learned conceptually. Because the NP and C groups were not significantly different, however, we cannot be certain that the difference between NP and TP students doesn't reflect selection bias.

The results of our analysis confirm that the NP curriculum reinforces student preferences for a self-directed learning environment. These students appear to adopt a learning behavior in which they conceptualized more and memorized less than their student counterparts in the TP. The patterns that surfaced among student responses reveal that the learning environment in which students are placed influences their learning behavior. The C students adapted to their learning environment and appeared to modify their views, preferences, and needs of their learning agenda over time, perhaps to resolve the dissonance between their initial preferences and what their learning environment required.

3. CLINICAL COMPETENCE:

A. REASONING AND DECISION-MAKING

Some of the faculty designers of the NP curriculum had

hypothesized that NP students would be more effective at clinical reasoning than their TP counterparts. Their rationale was that NP students, by discussing cases in their tutorials, were acquiring a knowledge base that was more relevant and closer in nature and structure to that formed by clinicians. The NP students were developing a repertoire of prototypical cases, reasoning through cases, and structuring their knowledge the way clinicians do (i.e., pattern recognition).

Although clinical reasoning lies at the center of the practice of medicine, its cognitive nature has remained elusive to investigators. No common view exists on the definition of clinical reasoning, the description of the process, or the criteria that distinguish strong clinical reasoning performances from weak. Thus, the dilemma faced by clinical faculty, house staff, and medical education researchers is: how does one pursue a viable examination of clinical reasoning when there exists so little agreement on its nature?

Our solution was to use methods of a variety of researchers as well as developing two of our own. We assessed students at two points in their curriculum: the end of their second year, before starting their major clerkships, and in their third and fourth years. The former testing consisted of a series of diagnostic evaluation work stations (see Appendix): the Laboratory Data Recall Task created by Geoff Norman; a modified oral exam designed by the three HMS medicine clerkship directors; the pattern recognition test; and a computer clinical case developed at Southern Illinois University (used for the 1989 class) or the Lab of Computer Science at Massachusetts General Hospital (the class of 1990). Clinical reasoning performance during the third year medicine clerkship was assessed by blinded analysis of the student's written clerkship evaluation by faculty and residents.

We found no differences between the group mean scores for any of the measures of diagnostic reasoning. Regardless of the type of instrument used to measure clinical reasoning (e.g., oral exam, recall task), the result in each situation was the same: no differences were detected in the ways NP, C, and TP students reason through a case.

B. CLINICAL PERFORMANCE:

To test clinical skills such as history taking, physical examination, and student-patient interaction, we used two approaches: content analysis of third year medical clerkship ratings and a work-station format with "standardized patients". In the clerkship ratings, a blinded reviewer content-analyzed faculty evaluation reports from the medicine clerkship for evidence of performance in the following areas: knowledge base; basic history and physical exam skills; clinical decision-making;motivation; and behavioral skills in relating to the patient. Because of the multiplicity of different raters in the many teaching institutions at Harvard, as well as the subjectivity and different language used by evaluators, we used the method of content-analysis of the

evaluations to make students' records more comparable. Medicine evaluations were chosen because medicine represents an integrative discipline. Although the rater was blinded to students' group membership, clerkship faculty generally knew which students were in the New Pathway.

The use of standardized patients for student assessment has gained rapid and widespread support among medical schools because of the work of Paula Stillman at University of Massachusetts, Emil Petrusa at University of Texas, and Howard Barrows at Southern Illinois University. Standardized patients were used for student assessment twice in this study: at the end of the second and fourth years. These patients were used to assess skills in interviewing and the student's style in relating to patients, as well as ability to elicit a history and make a diagnosis. Trained patients scored student performance and videotapes of the interaction were scored by two blinded interviewers using two rating scales developed to examine the quality of the doctor-patient interaction. In addition, students and standardized patients used the Barrett-Lennard Patient Relationship Inventory to rate empathic behavior in the interview.

There were no significant differences between NP and the other study groups in performance in the medical clerkship, either on overall grades or on content analysis subscores. In contrast, the NP students performed better on the interviews of the standardized patients; these differences are reported in detail in the next section..

4. PSYCHOSOCIAL DOMAIN

For the purposes of this investigation, we considered "psychosocial orientation" to have the following components:

basic knowledge in the behavioral sciences (e.g. understanding of common responses to illness, appreciation of major psychiatric disorders, awareness of the impact of psychological and social factors on health);

a set of skills in relating to patients that support the clinical functions of information gathering, development and maintenance of a therapeutic alliance (e.g. empathy, lack of judgmental qualities, warmth), and patient education;

attitudes and values towards the practice of medicine, the role of the physician and the determinants of health and illness that reflect a student-physician's basic predispositions and orientations in providing medical care.

The instruments used in this study are described in Appendix

I and II. The results of this investigation are summarized in Appendix IV.

Knowledge:

Students in both New Pathway classes scored significantly (p=.05) higher than the combined TP and C group on the Behavioral Science subtest of the National Board of Medical Examiners Part 1 examination. For technical reasons, we were able to do T-tests comparing NP with C students only for the class of 1990; they were higher (p=.04, one tailed test).

Attitudes:

The ATSIM, which showed that NP were more oriented than the C group towards preventive medicine at the beginning of medical school (p=.04), showed equalization of the preventive medicine focus between the two groups by the end of the first year. At the end of the second year, NP indicated a greater appreciation than did C students of the importance of factors affecting the doctor-patient relationship (p=.03). By the third and fourth years, however, all differences had disappeared, although our samples were considerably smaller.

There were no differences between the two groups on the dimension of locus of control during second or third year, measured by the Rotter Internal-External Locus of Control Scale, which is an indicator of the degree of control the individual feels s/he has over events in his/her life.

Differences in self-image among students in the three groups were demonstrated with the Q-Sort. In general, NP students (as compared with C) at the end of the second year described themselves as more psychosocially oriented (p=.02), more psychologically minded (p=.03), more emotionally expressive (p=.05), and more comfortable with ambiguity (p=.06) -- findings that we had hypothesized would characterize this group of students. There was also a trend for C students to characterize themselves as more protective towards others (p=.09), more cautious (p=.06), and more work-oriented (p=.10) than students in the NP at the end of the second year.

By the end of the fourth year, this configuration of differences had changed somewhat. NP described themselves as better able than C students to tolerate conflict (p=.003), more aware of personal limits (p=.000), more comfortable with emotional problems (p=.04), and more frustrated (p=.08), as well as less self-disciplined (p=.08). The NP students still characterized themselves as more psychosocially oriented, more emotionally expressive, and more psychologically minded, although these differences did not reach statistical significance in this small sample. These findings were amplified when we constructed scales linking similar items: NP students described themselves as more tolerant of uncertainty (p=.01), and less scientifically oriented (p=.07) than controls at the end of the fourth year.

At the end of the second year, NP students used the Q-Sort method to describe their image of the ideal physician -- a procedure which gives an indication of values about how physicians should be. In comparison to C, NP characterized the ideal physician as being more psychosocially oriented (p=.05), more introspective (p=.05), more depressed (p=.08), more self-blaming (p=.06), more aware of personal limits (p=.09), experiencing more congruence between personal and professional roles (p=.05), and as less intellectually confident (p=.06).

Evaluation of students' performance on the medicine clerkship revealed no significant differences in observed psychosocial orientation.

Skills:

Four independent measures were used to evaluate students' interviewing skills at the end of the second year. The Arizona Clinical Interview Rating Scale (ACIRS) showed that NP scored significantly better than C students on the entire instrument (p=.05). The Barrett-Lennard Patient Relationship Inventory, rated by the standardized patient, showed a consistent, but not statistically significant increment among NP students in comparison to C group. The New Pathway Interview Rating Form (NPIRF) was used by a blind rater to evaluate students' performance with patients. New Pathway students scored better than C on all 9 subscales of the instrument. Differences were statistically significant on 8 out of the 9 subscales, and the overall score of the NP students on the instrument was significantly higher than that of the C group (p=.01). The Interaction Analysis System for Interview Evaluation(ISIE) demonstrated that NP students were significantly more likely to focus on affective issues in the interview (p=.006), to use attentive silence (p=.001), and to allow the patient to talk more in the interview (p=.05).

All four instruments used to assess students' skills in relating to patients showed consistent and statistically significant differences in favor of the NP students. The confluence of these findings, using different methodologies that focus on divergent aspects of the interviewing and communication process suggests that NP students develop superior skills in relating to patients by the end of the second year.

By the end of the fourth year, many of the behavioral differences favoring the NP had diminished, although the samples are quite small. On both the ACIRS and the NPIRF, NP and C students show minimal differences; however, 30/45 of the individual items favored the NP students on the NPIRF; using the sign test, this finding is statistically significant at the p=.02 level. There were no significant differences between NP and C students on the Barrett-Lennard Patient Relationship Inventory. On the Ethics Problem Solving task, however, NP were significantly more likely than C and TP students to identify ethical problems (p=.02), and less likely to distance themselves from a patient with an ethical problem (p=.05).

Conclusions:

The design of this study allows us to differentiate self-selection effects from curriculum effects. The data reported above, comparing NP students to randomly selected Controls, whose entry characteristics were presumably similar to those of the NP students, suggest that the NP curriculum had a distinct, and positive impact on students' knowledge, attitudes, and skills in the first two years of medical school. The differences, however, attenuate in the third and fourth years, particularly in the area of interviewing and communication skills, but appear to be relatively preserved in the area of student attitudes. A measure of Ethical Problem Solving skills revealed important differences at the end of the fourth year, although further evaluation would be needed to definitively state that NP students' skills are superior to those of Controls alone.

It is not clear whether the apparent narrowing of differences between NP students and Controls during the clinical years represents relatively greater improvement by the Controls, worsening of skills among the NP students, or an artifact of the small and different fourth year sample. It might be argued that the clinical culture is so oriented towards biomedical issues and towards "efficiency", that more process-oriented interviewing behavior (attention to the patient's perspective, empathy, use of self as therapeutic instrument) would tend to decline, even among students who value a more humanistic approach. The persistence of attitudinal differences at the end of medical school, despite the powerful socializing effects of the hospital culture, suggests that curricular efforts to address attitudes (as occurred in the New Pathway) have the potential to have lasting effects. Follow-up studies to evaluate behaviors among trainees during residency will be useful in clarifying the implications, if any, of these attitudinal differences.

5. STUDENTS' PERSONAL EXPERIENCE IN THE DIFFERENT CURRICULA

We interviewed students at the end of the second and fourth years in order to assess the reactions of the three groups to medical school. These semi-structured interviews gave us information about the students' subjective impressions of the curriculum and their experience. In addition, we conducted a survey of fourth year students to develop more quantitative measures of their impressions.

Students in the two curricula reported quite different experiences during their preclinical years in medical school. The survey showed that NP were significantly ($p<.05$) more likely than the C group to describe their pre-clinical curriculum in the following terms: engaging, difficult, and useful. The C group were significantly more likely to select the following words to describe their curriculum: nonrelevant, passive, and boring. Both groups felt that they had great opportunity to develop close relationships

with their classmates, but the NP students rated themselves three times more likely than C students to have faculty role models and a close relationship with a faculty member during their preclinical years ($p<.001$). The NP group reported a higher degree of stress during their first two years ($p=.02$) but also scored somewhat higher on self-esteem during the first year. No differences in self-esteem between the groups emerged during years two, three and four.

Opinions about the clinical years were much the same in the NP and C groups. Both had similar positive opinions about the degree to which their preclinical curriculum prepared them for their clinical years. Close faculty relationships were equally likely for NP and C group students during years three and four.

Interviews largely corroborated the results of the survey. The analysis of the fourth year interviews has not yet been completed, but the second year interviews showed significant differences in the reactions of students to the two curricula. The traditional pathway students were fairly content without being gratified. The education they received at Harvard was close to what they were accustomed to, but without the pain of competing for grades to get into medical school. Many of them reported with surprise that the two years were much easier than expected, perhaps even easier than their last year or two of college. Even though many students expressed satisfaction on the one hand, they appeared, on the other hand, highly alienated. In fact many of the happiest students were those who had rejected the learning methods offered by the traditional curriculum and substituted their own. Conversely, some of the most unhappy students were those who were the least alienated, in the sense that they were at least trying to do everything, attend all the lectures and do all the reading. Many of them argued against the New Pathway, saying it was too much too fast and not structured enough.

On the other had, when asked to relate their positive educational experiences or to give suggestions to improve medical school, they often described New Pathway-like experiences.The lack of contact and occasional negative interactions with faculty were often cited as among the negative educational experiences.

What students from both programs liked is patient contact, the case-based approach with lots of time for small group discussion, having plenty of access to physicians as role models and mentors, learning clinical skills, and being able to observe physicians in action. The TP students also mentioned an appreciation of clarity, organization, and structure, both of individual teachers and of classes and the curriculum. Along the same lines, many students liked having camels (the verbatim lecture note syllabus).

There were two significant differences between the answers that NP and TP students gave to the question about positive educational experiences. First, a large number of NP students mentioned lessons about learning as their most memorable experience, and they seemed to find these important educational experiences in many more places and activities than did the TP students. Second, complaints about boredom were a constant refrain

among the TP students, while boredom was never mentioned by the New Pathway students.

Overall the NP students displayed a high level of engagement in their work. This engagement with the educational process had its negative side as well, often expressed as fairly high levels of stress, attributed to the following factors. First, students complained about lack of structure and organization of the overall curriculum. The problems included discontinuities in the curriculum, insufficient structure or guidance, and a poor sense of a master plan for it all. In these comments, students seem to reflect a belief that structured didactic learning is a necessary evil. They may not like memorization and long lecture time, but many assume it is necessary.

Second, tutorials were very stressful for many students. Many were upset by a group process that they felt reflected mainstream students and displayed lack of respect for diversity of values. This was particularly true for women, minorities and students coming from non-science backgrounds. Third, although the students seemed not to understand exactly how, they complained that being part of a controversial new curriculum resulted in additional tensions.

6. CAREER PLANNING

We examined two aspects of career planning to determine if there were differences between NP and C students: proportion of students taking time off during medical school (usually to do research or to take a second graduate degree) and choice of residency. The proportion of students taking time off was not significantly different for the two groups. Residency specialty choices were not significantly different between the two groups. A higher proportion of NP students chose primary care residencies (58%) than C (40%) or TP students (45%). However, this difference was not statistically significant.

SUMMARY AND CONCLUSIONS

The results of this study suggest that students in the new curriculum, when compared with a control group in the traditional medical curriculum, experienced a rather markedly different education, learned and employed more psychosocial skills, adopted different attitudes, and acquired a comparable level of knowledge and basic clinical skills. The NP students appeared to prefer a student-directed environment and to learn more for understanding, feel more challenged and less bored, and experience a greater degree of anxiety and stress than their counterparts in the traditional program. They knew, and were better known by, their faculty during the preclinical years. Their skills in relating to patients were more advanced, and their attitudes about the psychosocial aspects of their interactions more positive than

those of the control students. Despite undertaking a curriculum that placed more emphasis on and invested more time in social and behavioral aspects of medicine and that stressed self-directed learning, the NP students performed at least as well or better on virtually all measures of knowledge and clinical competence.

During the time of the FIPSE study, HMS was supporting a separate study of the students and faculty. In this ethnographic study, students were interviewed at least twice a year throughout medical school using methodological techniques developed by Byron Good and Mary Jo Delvecchio Good (1984). The interviews focused on two agendas, one set by the researchers and the other by the students, and yielded information on students' views on their medical education, their perception of the progression, their values and attitudes as reflected in their emerging role of the physician, their understanding and belief of the importance of competence and caring in medical practice; their view of and interest in research in medicine, and their sense of satisfaction with themselves and the world around them. This project's preliminary findings are quite consistent with those described in this report and will be reported later in a separate publication.

Putting these findings together, we believe that a credible argument can be made that the NP curriculum leads to a different type of studying and learning that reflects what adult learning theorists propound. This approach has important side benefits in encouraging contact and role modeling with faculty and the integration of biologic and psychosocial perspectives. Students perceive this to be a more difficult and challenging curriculum, generating more anxiety and stress about what and how much to study. Small group dynamics also appear to be stressful. On the other hand, these stresses are emblematic of challenges that must be mastered by successful clinicians during their professional socialization. This type of educational experience might socialize students differently to working in groups and to self-directed life long learning. However, this remains a theoretical perspective for now, since our study was not able to follow students long enough to determine if these exposures made a long-term difference.

This curriculum, despite widespread faculty concern that students might not learn what they need to know, appeared not to hamper cognitive or clinical skill performance. It is unlikely, since these conclusions are based on virtually full samples of the graduating classes, that we would have missed deficient performance in students undertaking this self-directed curriculum. These data should reassure schools whose student characteristics are generally similar to Harvard's that they are unlikely to damage the preparation of their students if they undertake such a curriculum.

The superior performance of students in psychosocial knowledge, skills, and attitudes is quite encouraging. Much criticism has been directed towards medical education's harmful effects on the doctor-patient relationship. This curriculum appeared to influence positively the students' skills in this area and their values about the importance of the relationship. Longer term studies will be needed to determine if these are enduring

changes, but these demonstrated results should encourage other schools to initiate similar interventions and suggest that educational process and milieux can make a difference in educational outcomes.

The comparability of the experimental and control groups makes the differences demonstrated in this study more meaningful. Selection bias has hampered and limited the interpretation of several prior curriculum evaluation efforts. Our experimental and control students were admitted to school without consideration for their competence to undertake the experimental curriculum and both groups wished to participate. The randomization process resulted in some slight differences between groups; these, however, are not sufficient to negate the demonstrated findings.

Other methodologic problems, however, may limit some of our conclusions. Because of poor student participation, much of the data from the clinical years is less reliable than that from the first two years of medical school. Selection bias and small numbers could have affected our results. This particularly affects the analysis of the simulated patient interactions and the clinical problem-solving. We cannot rule out the possibility that more significant differences in performance might exist at the end of four years, but we were unable to demonstrate them. On the other hand, the use of multiple instruments at different times made it possible for us to look for reinforcing findings. In general, the broad conclusions that we have drawn arise from the convergence of findings from more than one type of measure at more than one time, therefore lending credence to the findings. Some of our measures used secondary data (national boards, clerkship results, and residency choices) available for the whole class and are therefore not subject to such biases.

Generalizability of the evaluative data is always a concern to policy makers. The Harvard medical students are similar to those of the other dozen top schools in the US. The evidence from this study can probably be safely generalized to conclude that such a curriculum could be successfully used by students in these schools. Our conclusions are difficult to generalize to other schools whose students are less competitive. At least one other study that we have done (Moore, 1990), however, shows no difference in examination performance between students in the top and bottom quartiles of the Harvard class when they are in a problem-based curriculum. The experiences of a half dozen other schools with this approach, while not free of selection bias, has not suggested that the method creates significant problems among students who use it.

Faculty aptitude for such a new curriculum would also not be a major obstacle to its use. While being unique in many ways, this type of curriculum uses teaching methods mastered by faculty in many medical schools now. The number, variety, and level of teaching skills of the participating HMS faculty are achievable at most comparable US medical schools. Thus, we believe the data derived from this study are probably relevant to populations of students and to faculty at many, if not most, other schools.

Faculty members are often the best judge of the results of

educational innovations. We did not formally or systematically assess the opinions of our faculty about the new curriculum. However, the Harvard curriculum has progressively shifted to a more student-directed, problem-based format over the five years since the New Pathway began. Not all the faculty are enthusiastic about the new methods, but the majority of course leaders and department chairmen who have had direct experience with this method of teaching have come to favor such an approach in their courses. A number have changed from skeptics to enthusiastic supporters. Our faculty's reaction to this innovation provides strong indirect support for some of the positive outcomes demonstrated in this study.

Finally, some would argue that this type of curriculum does not warrant the extra effort needed to design and teach it. Our study did not study the cost differences between the NP and Traditional program. Very rough calculations suggest that the problem-based curriculum is, indeed, somewhat more expensive. However, since our study does demonstrate important educational benefits, this extra expense may well be warranted. Our findings lend concrete support to the arguments of educational theorists about the benefits of such an approach to teaching. Acquiring patterns of learning that prepare physicians for life-long study in a medical world without structured curricula is certainly a valuable outcome. Improvements in psychosocial attitudes and performance of their physicians will probably be worth a great deal to their future patients.

In conclusion, the evaluation activity of the New Pathway is entering a new phase at the medical school. A summary publication of the FIPSE project work is being prepared, as well as at least two detailed reports of the results in the psychosocial and learning behavior areas. These should be ready for submission for publication in refereed journals in 1991. We hope to seek grant funding to follow up the study groups near the end of their residency training programs. Many of the outcomes of the NP type of educational approach are likely to have longer term effects, and we would like to document if these actually occur. We are continuing to analyze some of the wealth of interview data that have been generated by the study. We expect that this will yield another report and probably a paper.

Finally, the Office for Educational Development at HMS is planning a national conference in June, 1991 in which the program evaluation will figure prominently. We will be presenting the methods and results during the conference, as well as organizing several workshops on evaluation methods in medical education.

References:

Association of American Medical Colleges. (1984). Physicians for the twenty-first century (Report of the Panel on the General Professional Education of the Physician and College Preparation for Medicine). Washington, D.C.: author.

Barrows, H.B. and Tamblyn, R.M. (1980). Problem-based learning. New York:
Springer.

Cronback, L.J. (1970). Essentials of Psychological Testing (3rd ed.). New York: Harper and Row.

Curry, L. (1983). Learning Style in Continuing Medical Education. Ottawa: Council on Medical Education, Canadian Medical Association.

Entwistle, N. (1981). Styles of Learning. Chichester: John Wiley and Sons.

Fox, R.D. (1984). Learning styles and instructional preferences in continuing education for health professionals: a validity study of the LSI. Adult Education Quarterly, 35, 72-85.

Freedman, R.D. and Stumpf, S.A. (1980). Learning Style Theory: less than meets the eye. Academy of Management Review, 5, 445-447.

Gallagher, J. (1978). The future of formal thought research: The study of analogy and metaphor. In B.F. Presseisen, D. Goldstein, and M.H. Appel (Eds.), Topics in Cognitive Development: Volume 2. New York: Plenum.

Gentner, D. and Stevens, A. (Eds). (1983) Mental Models. Hillsdale, N.J.: Erlbaum.

Good, M. and Good, B. (1984). Competence and Caring in the Professional Life Course of Physicians. Report of a research project, supported by the Arthur Vining Davis Foundation.

Institute of Medicine. (1983). Medical education and societal needs: A planning Report for the health professions (Publication No. IOM-83-02). Washington, D.C.: National Academy Press.

Joyhnson-Laird, P. (1983). Mental Models. Cambridge, Ma.: Harvard University Press.

Kosslyn, S. (1980). Image and Mind. Cambridge, Ma.: Harvard University Press.

Moore, G.T. The effect of compulsory participation of medical students in problem-based learning. In press. Medical Education.

Neufeld, V. and Norman, G. (Eds.). (1983). Assessing Clinical Competence. New York: Springer.

Newble, D.I. and Entwistle, N.J. (1986). Learning styles and approaches: Implications for medical education. Medical Education, 20, 162-175).

Painvin, C., Neufeld, V.R., and Norman, G. (1979, Nov). The "Triple Jump" exercise--a structured measure of problem solving and self-directed learning. Paper presented at the 18th Annual Conference of AAMC, Washington.

Piaget, J. (1971). Genetic Epistemology (E. Duckworth, trans.). New York: W.W. Norton.

Rappleye, W.C. (1932). Medical Education (Final Report of the Commission on Medical Education). New York: Association of American Medical Colleges' Commission on Medical Education.

Rumelhart, D.E. (1980). Schemata: The building blocks of cognition. In R.J. Spiro, B.C. Bruce, and W.F. Brewer (Eds.), Theoretical Issues in Reading Comprehension. Hillsdale, N.J.:Erlbaum.

Sachs, L.A. (1985). A comparison of medical students' performance in independent study and traditional programs. Journal of Medical Education, 60, 602-609.

Schmidt, H.G., Dauphinee, W.D., and Patel, V.L. (in press). Comparing the effects of problem-based and conventional curricula in an international sample. Journal of Medical Education.

Smith, F. (1975). Comprehension and Learning. New York: Holt, Rinehart, and Winston.

West, D.A., Umland, B.E., and Lucero, S.M. (1985). Evaluating student performance. In A. Kaufman (Ed.), Implementing Problem-Based Medical Education. New York: Springer.

Wijnen, W. and Bouhuijs, P. (1986, October). The Maastricht Progress Test: An example of comprehensive knowledge testing. Paper presented at the 97th Annual Conference of AAMC, New Orleans.

from Harvard project 1987-9

APPENDIX:

I. EVALUATION INSTRUMENTS:

The following briefly describes the instruments used in the evaluation of the New Pathway curriculum. The abbreviated identifier that we use to name each instrument in Table A is given in parentheses.

Biographical Data (Bio Data): Demographic information about the student's background, including family education, geographic location, college and research experiences.

Career Plans (Career Plans): Taken from Tarlov, U. of Chicago, National Study of Internal Medicine Manpower). Student's expected career plans for specialization or type of work experience.

Career Expectations (Career Exp): Taken from Tarlov. A series of questions about the student's expectations of the practice of medicine, including public attitudes, availability of funding, and expected work life of the physician.

Personality Research Form (Pers Res Form): developed by Jackson (Research Psychologists Press). The PRF is a 64 item questionnaire eliciting descriptions of personality and attitudes about interpersonal interactions.

Preferred Learning Style Index (PLSI): developed by Howard Stone at the University of Wisconsin Medical School as one of the mechanisms to monitor Wisconsin's own innovative medical curriculum. The survey measures a student's preference for either a structured or unstructured learning environment. A structured environment is defined as one in which students are passive receivers of information: they are given their educational objectives, told what they should learn, and observe demonstrations of material and skills to be learned. In an unstructured environment, students are active learners: they are expected to be personally involved in setting learning objectives, actively involved in discussion, and responsible for monitoring their own learning.

The PLSI yields two scores: a discovery score and a receptive score. The discovery score reflects a student's preference for an unstructured environment; the receptive score reflects a student's preference for a structured environment. We created a third score, the PLSI score, which is simply the difference between the other two. The higher the PLSI score, the stronger the preference for an unstructured environment, and the lower the score, the stronger the preference for a structured environment.

Learning Style Index (LSI). The LSI examines the process of learning along two dimensions: the degree of preference that a learner expresses for abstract over concrete experiences and for action over reflection. Based on learners' preferences, they are classified into one of four learning styles: accommodators, assimilators, divergers, convergers.

Short Inventory of Approaches to Studying (Entwistle):
Developed by Noel Entwistle in collaboration with European and Australian colleagues, this instrument examines learning behavior of students in terms of the depth (i.e., surface versus deep structure) of understanding that students achieve by their learning style.(Entwistle, N. (1981). Styles of Learning. Chichester: John Wiley and Sons. Newble, D.I. and Entwistle, N.J. (1986). Learning styles and approaches: Implications for medical education. Medical Education, 20, 162-175).

Rotter Internal-External Locus of Control Scale -Revised (Locus of Control): This forced choice questionnaire is designed to evaluate the degree of control the individual feels over events in his/her life. A higher score (external locus of control) reflects a more fatalistic viewpoint; a lower score (internal locus of control) indicates a greater belief in one's own ability to affect events and to control one's life (Lefcourt HM (1976) Locus of Control: Current Trends in Theory and Research. Lawrence Erlbaum Associates, Hillsdale, NJ.Rotter, J.B. (1966). Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 80, (Whole No. 609).

Attitudes Towards Social Issues in Medicine (ATSIM):
developed by Drs. Parlow and Rothman of the University of Toronto. The ATSIM measures students' attitudes about social factors in medicine, preventive medicine, and doctor-patient relations. Students who score high: on the social factors scale appear to recognize the importance of social factors in the patient's environment as determinants of his health or illness; on the preventive medicine scale appear to recognize the role of preventive medicine in maintaining health, and seem willing to adopt the methods this implies; and on the doctor-patient relations scale appear to recognize the importance of emotional factors and interpersonal relations between physician and patient in effective patient care.

Work Environment Scale (Work Environment):
The Work Environment Scale was adapted for use in medical education by Dr. Roy Feldman from an instrument developed by Rudolf Moos. The WES has typically been used by psychologists to obtain employees' perceptions of their work environment. Our adaptation -- the HMS Environment Scale -- was developed to yield the same measures as Moos' original instrument: involvement with and concern about HMS goals; importance of peer cohesion; desire for supportive faculty; desire for autonomy;preference for environment that is task oriented; preference for high work pressure; desire for clarity and explicit directions; preference for faculty control of learning; preference for innovation; degree of importance of physical comfort of learning environment.
(Moos, R. (1974). Evaluating treatment environments: a social ecological approach. N.Y.: Wiley).

Cognitive Behavior Survey (Cognitive Survey): developed by Rudolph Mitchell, HMS. The survey was constructed as a vehicle for

exploring the nature of students' learning behavior as it develops in an academic medical environment. The survey primarily measures the extent to which memorization and conceptualization behaviors are present in a student's learning style. Memorization refers to cognitive processes that center on rote learning (e.g., drill and repetition, use of flash cards). Conceptualization relates to those active processes that occur in the construction of mental models, the visualization of physiological processes, or the clear, concise summarization of complex material. The survey contains a set of seven-point differential items, subsets of which comprise a memorization scale and a conceptualization scale. Student responses result in two scores: a memorization score, which is the raw score total for the memorization items; and a conceptualization score, which is the raw score total for the conceptualization items.

National Board Exam, Parts I and II (National Boards I,II): Almost all medical students in the U.S. take this exam at the end of their second and fourth years. Part I is an objective, standardized multiple-choice exam designed by the National Board of Medical Examiners that tests students' knowledge in seven areas: anatomy, biochemistry, microbiology, pathology, physiology, pharmacology, and the behavioral sciences. Currently, the NBME represents the best standardized measure of what the medical community believes the knowledge base of medicine should be.

Pattern Recognition Test (Pattern Recog): an experimental component of the NBME, the pattern recognition test is intended for use in future versions of part II of the NBME. This test asked students to identify a diagnosis associated with a two or three item set of descriptive characteristics. This test has shown good psychometric characteristics in its early assessment by the Boards. Recent cognitive psychology research suggests that physicians structure their knowledge bases around profiles of signs and symptoms in which diseases present themselves. The pattern recognition test was thus viewed as an additional measure of content knowledge.

Biochemistry Review Exam (Biochem): developed by Manfred Karnovsky at HMS. A special essay examination given to fourth year students to assess their understanding of the biochemical aspects of frequently encountered clinical problems.

Laboratory Data Recall Task (Lab Recall): developed by Geoff Norman, McMaster University. In this test, students are given a set of brief written scenarios each of which details chief complaints of a patient and the accompanying laboratory values. Students are then asked to generate a diagnosis for each case and to recall as many lab values as possible. This recall task is a component of the research carried out by Geoffrey Norman (McMaster University) on expert-novice differences among clinicians. The task measures the extent to which learners organize and recall information in the form of clinical patterns. Norman had demonstrated that the recall of lab values is a function of the extent to which learners have stored their clinical knowledge as

the patterns in which the disorders clinically present themselves. Those with a stronger clinical context, like more advanced clinicians, were expected to be able to recall discrete data items more accurately and completely than those with less clinical sophistication.

Structured Oral Examination (Oral Exam): developed by Kenneth Falchuck, Leslie Fang, Mark Peppercorn, and Gordon Moore at HMS: In the structured oral interview, the clinician presents a summary of a history and physical of a patient admitted to the Emergency Room. The student's task is to describe the sequence of actions s/he would perform, the lab tests s/he would order, and the management actions s/he would take. With each exam procedure or lab test the student names, the clinician records on a blackboard the appropriate finding or lab value. Once the student has gathered all the data s/he needs and has taken all of the necessary emergency actions, the student presents a problem list and the corresponding differential. The student then reasons through the data and reaches a final diagnosis. Throughout the procedure, the clinician asks the student to justify his/her proposed actions and hypotheses. The model is structured to allow the clinician clear access to the student's knowledge base and level of clinical reasoning. The model for the particular case in question includes at least three or four scenarios, any one of which may be played out by the clinician depending on the work-up and interventions ordered by the student. At any time the clinician can transform the emerging scenario into one of the alternatives in order to prolong the student's diagnostic pursuit. This provides the clinician with numerous situations in which to construct an in depth evaluation of the student's cognitive performance in a clinical setting. Based on his/her observations and student responses during the interaction, the clinician scores the quality of student's clinical reasoning in a number of areas including: the appropriateness of the student's data collection; the effective use of the data; and the line of reasoning employed in generating, confirming, and refuting working hypotheses about the differential diagnosis. A total score is computed from the sum of all subscores.

This approach was selected to serve as the basis for a new examination because other structured examination based on similar methods have been shown to be effective assessment mechanisms (Painvin, C., Neufeld, V.R., and Norman, G. (1979, Nov). The "Triple Jump" exercise--a structured measure of problem solving and self-directed learning. Paper presented at the 18th Annual Conference of AAMC, Washington. West, D.A., Umland, B.E., and Lucero, S.M. (1985). Evaluating student performance. In A. Kaufman (Ed.), Implementing Problem- Based Medical Education. New York: Springer).

Southern Illinois University Computer Case (SIU case): developed by Howard Barrows, Nu Viet Vu, and Reed Williams. The Southern Illinois University (SIU) computer case is a computer-based assessment of diagnostic reasoning that monitors students' thinking in their gathering of data from a patient's history,

physical exam, and lab tests in order to formulate a diagnosis. Students collect patient information by asking the computer for specific information. As they work through the case, students are required to link the data they obtain to the different diagnoses they are considering as part of their differential.

MGH Computer Assessment Program for Clinical Reasoning (MGH Case): A computer-based diagnostic assessment using randomly assigned clinical scenarios. Students are allowed to solicit history, physical exam, and laboratory data in their pursuit of a diagnosis. Scoring is based on appropriateness of data collection and on accuracy of final diagnosis.

Health promotion/ disease prevention test(HP/DP Test): This instrument was designed to elicit student's response to a clinical vignette that contained several questions related to HP/DP (evaluation of risk, counselling, predictive validity of clinical tests). Developed at HMS by William Taylor and Rick Pels.

History-taking Task on Standardized Patient: Standardized patients are individuals trained by Paula Stillman of the University of Massachusetts (the development of this methodology was supported by a FIPSE award) to function as patients by giving histories and portraying various physical disorders, e.g. diminished breath sounds or elevated blood pressure, that are consistent with real situations. These portrayals are standardized and can be replicated for each student examined. In an assessment that uses standardized patients, each student rotates through a series of stations and performs specific tasks on a standardized patient at each station. These tasks vary from taking a history or performing specific aspects of a physical exam to relating distressing news to a patient about his/her child.

The use of standardized patients for student assessment has gained rapid and widespread support among medical schools because of the work of Paula Stillman at University of Massachusetts, Emil Petrusa at University of Texas, and Howard Barrows at Southern Illinois University.

History-taking Task on Standardized Patients: Arizona Clinical Interview Rating Scale (ACIRS): This instrument, developed by Paula Stillman, is widely used for student assessment in conjunction with the use of standardized patients. The ACIRS is a rating form used by the standardized patients to evaluate student performance. The rationale for the use of standardized patients, and the development and use of the ACIRS are discussed in (Stillman PL et al. (1986). Assessing clinical skills of residents with standardized patients. Annals Int Medicine. 105:762-771).

History-taking Task on Standardized Patient:New Pathway Interview Rating Form (NPIRF): This 65 item instrument, rated on a 1-5 Likert scale (5=excellent; 1=poor), was designed to reflect the behavioral objectives of the New Pathway curriculum, and consists of 9 subscales: 5 "technical" skills (Opening the Interview, Technical Questioning Skills, Organization of the Interview,

Patient Education, Closing the Interview), and 4 "interpersonal skills" (Use of Self as Therapeutic Instrument, Empathy, Attention to Patient's Perspective and Non-verbal Behaviors).Developed by Susan Block MD at HMS.

History-taking Task on Standardized Patient:Interaction Analysis System for Interview Evaluation(ISIE): Designed by Bryce Templeton This is a reliable and well-studied quantitative instrument that relies on microanalysis of doctor-patient interactions to codify specific behaviors in the medical interview. It produces a quantitative description of interviewer behaviors and patient responses, reflecting a variety of parameters of interviewing behavior.(Templeton B, MacDonald M. (1982) Use of interaction analysis i assessing physician trainee interpersonal skills. in Lloyd JS (ed) Evaluation of Noncognitive Skills and Clinical Performance. Chicago, American Board of Medical Specialties, 155-167).

Barrett-Lennard Patient Relationship Inventory (Barrett-Leonard): This instrument has been widely used in research on therapist-patient interactions, mostly in mental health settings. We selected it for our study to allow us to deepen our appreciation of differences in empathic responsiveness among the student groups. The scale generates two global ratings -- one on empathy -- the student's capacity to vicariously understand the patient's experience -- and one on regard -- the student's overall stance of esteem, respect, and appreciation of the patient.

Hogan Empathy Scale (Hogan): This instrument, widely used in medical education settings, was designed to assess empathic orientation and was used to evaluate students' empathic abilities when they entered medical school. Reference: Greif EB, Hogan R (1973). The theory and measurement of empathy. J Consulting Psychology.20;3: 280-284.

Q-sort Task (Q-sort): developed by Susan Block at HMS to determine self-perceptions and perceptions of the "ideal" physician. The Q-Sort methodology is an approach to psychological measurement that has been widely used to study personality development. Each student was asked to rank order a set of 72 descriptive phrases that were culled from the medical education literature and the attitudinal objectives of the NP. Students rank ordered the descriptive items in describing themselves and their image of the ideal physician. (Block, J. The Q-Sort Method in Personality Assessment and Psychiatric Research.)

Ethical Problem Solving Task (Bioethics): Building on the work of Spooner et. al., this measure presented students with a common medical ethics dilemma on videotape, and asked them to: identify the important ethical issues in the vignette, describe their emotional reactions to the situation, define additional information that would be needed to decide on a course of action, and commit to a series of actions in response to the situation. Student responses were blindly scored using a set of consensually-generated

criterion responses. Reference: Spooner HJ, Haight KR, Emson HE, To T (1989). Assessment of medical students' learning and performance in an introductory medical ethics course. Teaching and Learning in Medicine. 1:3:167-170.

Measure of Intellectual Development (MID): Developed by Knefelcamp and Widick at the University of Maryland based upon work of Perry. The MID (Measure of Intellectual Development) is based on Harvard psychologist William Perry's work characterizing the stages of intellectual growth of the individual. Learners at the lowest stage view knowledge as absolute, dualistic in nature, and a compilation of facts. At this stage, learners do not question the knowledge that they receive but instead accept it without critical examination. In contrast, the learners at the highest stage view knowledge in a relative context, not in an absolute framework. They analyze and synthesize their knowledge. They critically examine new knowledge and view their professors as resources, not as authoritarian figures who transmit perfect knowledge to their students.

The MID requires students to write an essay in which they describe in detail the best learning experience they have encountered. Trained specialists score the essays in terms of the Perry schema of intellectual development. The score assigned to an essay represents the student's stage of intellectual development as reflected by his/her responses.

Defining Issues Test (DIT). The DIT is a paper and pencil test that was designed by James Rest of University of Minnesota based on Lawrence Kohlberg's work on moral reasoning. It is the most frequently used instrument to assess moral reasoning in medical education (Schlaefle, A., Rest, J.R., and Thoma, S.J. (1985). Does moral education improve moral judgement? A meta-analysis of intervention studies using the Defining Issues Test. Review of Educational Research, 55, 319-352).

Medical Clerkship Evaluation (Med Clerk): We evaluated clinical reasoning ability, clinical skills, motivation, fund of knowledge, and psychosocial orientation of all the study groups during their required medicine clerkship, usually taken in the student's third year. Written assessments, prepared from evaluative comments of faculty and residents, are available for all students who have completed the clerkship. These assessments were content analyzed by a blinded reviewer for these parameters of performance. Numerical performance scores were given to each student on a scale of one to five.

Second Year Exit Interview (Yr2 Interview): Developed and administered by Carolyn Briggs-Style at HMS. A semistructured interview about the experience of the first two years of medical school.

Fourth Year Exit Interview (Yr4 Interview): Developed and administered by Carola Eisenberg at HMS. A semistructured interview about the experience of the all four years of medical school.

Fourth year exit survey (Yr4 Survey): developed by Gordon Moore at HMS. A survey eliciting student opinions about the most important attributes of their medical school curriculum and educational environment.

Internship Listing (Internship): used to determine career choice.

II. TABLE OF TESTS

EVALUATION INSTRUMENT	CLASS	YEAR GIVEN				AREA OF PERFORMANCE MEASURED BY INSTRUMENT
		1	2	3	4	
BIO DATA	89/90	*				Initial comparability of NP and C groups
CAREER PLANS	89/90	*	*		*	Initial comparability of NP and C groups. Determine changes in and differences between NP and C groups over time.
PERS RES FORM	90	*				Initial comparability of NP and C groups.
CAREER EXP	89/90	*	*		*	Initial comparability of NP and C groups. Determine changes in attitudes over time.
PLSI	89/90	*	*			Assess initial comparability of learning preferences. Changes over 1st year.
LSI	90		*			Differences in learning styles.
ENTWISTLE	89	*				Initial comparability in learning styles.
LOCUS OF CONTROL	89			*		Differences in sense of internal or external control over student's work in medical school. Learning style.
	90		*	*		
ATSIM	89/90	*	*	*	*	Initial comparability and changes in attitudes towards social and preventive medicine over time. Psychosocial skills
WORK ENVIRONMENT	89		*	*		Changes in preference for characteristics of learning environment over time.
	90	*	*	*		
COGNITIVE SURVEY	89		*		*	Differences in approaches to learning.
	90		*			
NATIONAL BOARDS	89/90		*		*	Differences in knowledge base.

EVALUATION INSTRUMENT	CLASS	YEAR GIVEN				AREA OF PERFORMANCE MEASURED BY INSTRUMENT
		1	2	3	4	
PATTERN RECOG	90		*			Content knowledge, clinical reasoning
BIOCHEM	89				*	Content knowledge retention
LAB RECALL	89		*			Cognitive performance:content knowledge
ORAL EXAM	89		*			Content knowledge and clinical reasoning
SIU CASE	89		*			Content knowledge and clinical reasoning
MGH CASE	90		*			Content knowledge and clinical reasoning
HP/DP TEST	90		*			Content knowledge and clinical skills
ACIRS (STANDARDIZED PATIENT)	89/90		*		*	Content knowledge and clinical reasoning; Psychosocial and interviewing skills
NPIRF (STANDARDIZED PATIENT	89/90		*		*	Psychosocial and interviewing skills
ISIE (STANDARDIZED PATIENT)	89/90		*		*	Psychosocial and interviewing skills
BARRETT-LEONARD	90				*	Psychosocial (empathic behavior)
HOGAN	89	*				Initial comparability of NP and C across a measure of empathy
Q-SORT	89/90		*		*	Psychosocial (attitudes about themselves and the ideal physician)

EVALUATION INSTRUMENT	CLASS	YEAR GIVEN				AREA OF PERFORMANCE MEASURED BY INSTRUMENT
		1	2	3	4	
BIOETHICS	89/90				*	Psychosocial (ethics)
MID	89				*	Cognitive: Learning behavior and intellectual development
	90		*		*	
DIT	90	*				Entry characteristics:psychosocial (ethical development)
MED CLERKSHIP	89/90				*	Clinical skills; motivation; cognitive (knowledge base); cognitive (decision making); psychosocial orientation.
YR2 INTERVIEW	90		*			Experience of the first two years of medical school
YR4 INTERVIEW	89/90				*	Experience of all four years of medical school
YR4 SURVEY	89/90				*	Experience of all four years of medical school curriculum
INTERNSHIPS	89/90				*	Career choice

III. LEARNING BEHAVIORS SUMMARY TABLE

LEARNING BEHAVIOR (Preclinical Years)[1] Summary of Survey Results		
SURVEY	'89	'90
PLSI		
First Year	*	*
Second Year	*	*
Entwistle's Short Inventory		not administered
Kolb's LSI		not administered
Cognitive Behavior Survey	*	insufficient return
MID (2nd year format)	not administered	**
* indicates significant differences at p.≤05 between NP and C groups ** indicates significant difference at p=.01 between NP and TP groups		

[1] Insufficient sampling of fourth year students prevented any statistical analysis of the PLSI and MID for the class of '90 and of the Survey of Cognitive Behavior in a clinical setting for the class of '89.

IV. PSYCHOSOCIAL SUMMARY TABLE

OVERVIEW: PSYCHOSOCIAL DOMAIN: RESULTS

CATEGORY/TEST	YEAR	RESULT	P.=
KNOWLEDGE			
NBME Pt.1 Behavioral Sci.	2	89: NP>T+C	.05
		90: NP>C	.05★
NBME Pt.2 Behavioral Sci.	4		
ATTITUDES			
Rotter Locus of Control	2	NP≈C internal vs. external control	
	3	NP≈C internal vs. external control	
ATSIM	1	NP>C Preventive Med NP≈C Dr/Pt Relation	.04
	2	NP≈C Preventive Med NP>C Dr/Pt Relation	.03
	3	NP≈C Preventive Med NP≈C Dr/Pt Relation NP≈C Social Factors	
	4	NP≈C Preventive Med NP≈C Dr/Pt Relation NP≈C Social Factors	
Q-Sort: Self	2	NP>C: psychosocially oriented psychologically minded emotionally expressive comfortable with ambiguity NP<C: protective towards others cautious work-oriented	 .02★ .03★ .05★ .06★ .09 .06 .10

CATEGORY/TEST	YEAR	RESULT	P.=
	4	NP>C: able to tolerate conflict tolerant of ambiguity aware of personal limits comfortable with emotional problems frustrated NP<C: self-disciplined scientifically oriented personal/professional selves congruent	 .003★ .01 .000★ .04★ .08 .08 .07 .05
Q-Sort: Ideal	2	NP>C: psychosocially oriented introspective depressed self-blaming aware of personal limits intellectually confident personal/professional selves congruent	 .05★ .05 .08 .09 .09 .06 .05
Clerkship Assessment: Psychosoc. Orientation	3	NP≈C	
SKILLS			
Hogan's Empathy	1		
History: Information	2		
	4		
History: ACIRS	2	NP>C	.05
	4	NP≈C>T	
History: Barrett-Lennard	2	NP>C,NP>T	NS
	4	NP≈C	
History: NPIRF	2	NP>C on 9/9 scales and overall NP>C (overall using sign test)	.01 .000

CATEGORY/TEST	YEAR	RESULT	P.=
	4	NP≈C (T-Test) NP>C (overall using sign test)	 .02
History: ISIE	2	NP>C: affective focus attentive silence Patient talk	 .006 .001 .05
Ethical Problem Solving	4	NP>C: identification of ethical problems NP>C+T: identification of ethical problems NP<C+T: distance selves emotionally from pt	 NS .02★ .05★

★ = one-tailed test of significance used

APPENDIX FOR FIPSE:

We found FIPSE and our project officer to be most supportive during the course of the Grant. Interested and concerned, Connie Cook followed the course of the study up until her departure from FIPSE and offered helpful suggestions when appropriate. The FIPSE annual meetings were interesting to us but not terribly relevant to our field of study.

We believe that FIPSE could continue to play an important role in medical education in the future. We would suggest that FIPSE look for that unusual opportunity where an important educational problem or approach, an unusual chance to demonstrate or evaluate it, and an influential school occur together. Our experience suggests that an investment in that circumstance can yield important information and leverage on change. We have continued to be impressed with the number of schools that have been interested in the evaluation of the New Pathway and who ask for information, come to visit, or attend visitor days at Harvard Medical School. We think the publication of the report will enhance this interest and encourage other schools to change their way of teaching.